Performance Enhancements in a Weather Prediction and Forecasting System Using Deep Learning Techniques

Synopsis of the Research Proposal

Submitted for the Registration in the

Ph.D. Program (Part-Time) in Engineering

By

Lakhan Bhaskar Kadel

Enrollment Number:*23RDCDIUJDD02011*

Under Supervision of

Dr. Mukesh Kalla

Associate Professor(CSE),Head(CSE),Faculty of Science and Technology

THE ICFAI UNIVERSITY JAIPUR

LIST OF CONTENTS

Abstract	**3**
1 Introduction	**4-11**
2 Literature Review	**12-29**
3 Research Gap	**30**
4 Objective	**31**
5 Methodology	**32-35**
6 Work Plan for the Research	**36**
7 References	**37-44**

ABSTRACT

Weather prediction remains one of the most challenging yet crucial tasks in environmental science, with far-reaching implications for agriculture, transportation, disaster preparedness, and public safety. This research presents an innovative weather forecasting system that integrates multiple machine learning paradigms with modern web-based deployment architecture.

This research focuses on the development of an enhanced weather forecasting system that integrates machine learning (ML) and deep learning (DL) techniques to improve the accuracy and reliability of weather predictions. The primary objective is to address the challenges in weather prediction by leveraging the capabilities of ML and DL for more precise forecasting of meteorological parameters such as temperature, atmospheric/surface pressure, humidity, precipitation, and wind speed. The study begins with a comprehensive review of the key challenges faced in weather prediction and the application of machine learning and deep learning techniques for forecasting. A weather prediction system is then designed and developed that combines both ML and DL approaches, enabling more accurate and efficient forecasts. The research further analyzes historical weather data to identify patterns and trends, using various ML and DL models to uncover insights and improve forecasting capabilities. The performance of the developed system is compared to traditional weather forecasting models, emphasizing accuracy and computational efficiency. Finally, different ML and DL algorithms are evaluated and optimized for weather forecasting tasks, demonstrating the potential of these advanced methods to revolutionize weather prediction and enhance decision-making across various sectors.

1. INTRODUCTION

1.1. Introductory Background

Weather Forecasting or prediction is the scientific approach of estimating the climatic conditions at a given location and time, according to the interpretation of weather data. This information is collected from multiple sources, such as satellites, radars systems, terrestrial stations, weather balloons, and sea-based buoys, which measure weather factors like temperature, pressure, moisture, wind velocity and rainfall [1], [2].

Weather Forecasting holds a key position in today's modern society, offering significant insights that affect various industries, including farming, transportation, power, disaster management and daily activities [3], [4], [5]. Accurate weather predictions guide farmers in planning irrigation, sowing, and harvesting, minimizing the effects of harmful weather conditions [6], [7]. They are also fundamental to air and sea transportation sectors, where safety is highly depended on understanding the atmospheric conditions.

Additionally, weather forecasting plays a crucial role in disaster management systems, delivering advance alerts for severe weather conditions like hurricanes, cyclones, tornadoes and floods. These alerts allow authorities to execute early evacuation strategies and emergency actions, ultimately protecting lives and minimizing the financial and economic losses [8], [9].

The development of cutting-edge technologies such as satellite imaging, remote sensing, powerful computing systems has greatly improved the accuracy and reliability of weather predictions [10]. In addition, the use of machine learning and artificial intelligence has enhanced forecasting methods, allowing meteorologists to process large amount of datasets and detect complex patterns in weather behaviour [11].

Although weather forecasting is complex and challenging, it is continuously improving and becoming more accurate and easier. From giving daily weather reports to enabling long-term climate research, it helps individuals, organizations, corporations and authorities to make knowledgeable decisions and respond to the unpredictable patterns of Earth's weather [12], [13].

In short, weather forecasting is not only a scientific field but also a crucial tool for enhancing adaptability and planning in a world that is becoming more unpredictable.

1.2. Types of Weather Forecasting and Prediction

Weather forecasting is classified into various types based on the time frame and scope of predictions. Each type serves distinct purposes, ranging from short-term weather alerts to long-term climate analysis. The following sections explore the major types of weather forecasting in detail.

1.2.1. Nowcasting

Nowcasting is the practice of predicting short-term weather conditions, typically for duration of up to six hours. It targets on specific regions and offer accuracy, real-time data on rapid changes in the atmospheric conditions. This type of weather forecasting methods uses high-quality data from Doppler radars, satellites and ground monitoring systems to monitor fast weather changes like thunderstorms, fog, or sudden drop in the temperature [14].

Weather experts use advanced tools and algorithms to study the changing behaviour of weather systems during nowcasting. As an example, radar images are used to track rainfall levels and predicting the storm movement, while satellite imagery offers updates on the formation and breakdown of clouds [15].

Key Points:

- Time Window: A span of 6 hours.
- Perfect for tracking temporary or short-duration weather events.
- Frequently used in aviation, urban development, and event planning.

1.2.2. Short-Range Forecasting

Short-range forecasting involves predicting weather for next 1 to 3 days. It offers basic weather information, including temperature, precipitation, wind velocity and humidity, which are important for daily activities and short-term decision-making. This type of weather prediction combines observational data with mathematical models to generate relatively accurate and dependable predictions. Short-term weather forecasts play a key role in various fields. Farmers mainly rely on them to plan irrigation and protect crops from adverse weather conditions. In a similar way, they are essential for the logistics and transportation sectors, guaranteeing timely and secure operations [16], [17].

Key Points:

- Duration: 1 to 3 days.
- Depends on observational data and computational models.
- Essential for farming, logistics, and outdoor activities.

1.2.3. Medium-Range Forecasting

Medium-range forecasting expands the forecast window to 3 to 7 days, concentrating on larger atmospheric trends and patterns. These models basically use advanced mathematical formulas to represent the interactions between atmospheric factors. As the prediction period increases, the accuracy of medium-range forecasts declines, because of the unpredictable nature of the atmosphere. Despite this, they continue to be widely used for planning activities, such as organizing outdoor events, managing energy resources, and coordinating shipping operations [18], [19].

Key Points:

- Forecast Range: 3 to 7 days.
- Uses advanced numerical simulations
- Beneficial for mid-term strategies across various sectors.

1.2.4. Long-Range Forecasting

Long-range weather predictions focus on conditions expected over weeks to several months. The focus is on detecting trends or patterns such as seasonal temperature fluctuations, rainfall variability, or drought conditions. Long-range weather predictions rely on statistical modelling and historical datasets, frequently enhanced by global climate model outputs. Such forecasts play a crucial role in agriculture, water resources, and energy sectors by shedding light on long-term climatic patterns. For example, they enable farmers to plan for seasonal droughts or floods [20], [21].

Key Points:

- Prediction Range: From Weeks to months.
- Highlights general trends and patterns instead of minute details.
- Provides guidance for seasonal preparation and resource optimization.

1.2.5. Climatic Prediction

Climatic prediction involves studying long-term trends and averages in atmospheric behavior over decades and centuries. It is utilized for examining the climate change, global warming, and broad scale environmental patterns. These climate predictions are based on a wide range of datasets, including historical climate records, oceanic temperature profiles, and atmospheric gas levels. Climatic predictions are essential for guiding policy based decisions and planning for long-term sustainable development. These predictions provide insights for developing strategies to reduce the impact of climate change effects, like reducing greenhouse gas discharge, modifying infrastructure and protecting the natural ecosystems [22], [23].

Key Points:

- Forecast period: Decades to centuries.
- Focuses on long-term climate shifts and atmospheric trends.
- Fundamental in addressing global environmental concerns.

1.3. Techniques Used in Weather Forecasting

Weather forecasting relies on a variety of scientific and computational techniques to ensure accurate and timely predictions. These methods range from traditional statistical models to advanced machine learning and dynamic simulations. The following sections outline the most widely used forecasting techniques and their applications.

1.3.1. Numerical Weather Prediction (NWP) Models

Numerical Weather Prediction (NWP) models form the foundation of today's modern weather forecasting. These models utilizes mathematical equations based on the physical laws like thermodynamics, fluid dynamics and mass-energy conservation to represent the atmospheric behavior. Through the integration of current atmospheric data into these models, allows the NWP model to predict the future weather patterns within a defined time period [24].

Advanced computing systems are required to execute these models because of the vast complexity and volume of the datasets. High-level NWP models like Global Forecast System (GFS) and the European Centre for Medium-Range Weather Forecasts (ECMWF) are widely trusted for their accuracy and are also used globally by many meteorologists agencies. These

models deliver weather predictions across multiple time intervals, from hours to weeks, and they are especially useful for medium to long-term predictions [25].

Key Points:

- Founded on established physical laws (thermodynamics, fluid dynamics).
- Examples: GFS, ECMWF, Weather Research and Forecasting (WRF).
- Extremely accurate, but requires substantial computing resources.
- Particularly useful for medium to long-term weather predictions.

1.3.2. Machine Learning Algorithms

The integration of Machine Learning has completely reshaped weather forecasting, which allows the analysis of large and complex datasets. These algorithms continuously learn from past weather events and refine their predictions over time. When compared to NWP models, machine learning is capable of identifying non-linear correlations and hidden trends in the data, providing significant benefits for short-term and localized weather predictions [26], [27], [28].

Some commonly used machine learning algorithms in weather forecasting include:

- **Support Vector Machines (SVM):** Categorizes weather conditions, such as storm or clear skies.
- **Random Forests:** Predicts distinct results like rain or no rain by constructing several decision trees and combining their predictions.
- **Neural Networks:** Copy human brain functions to capture complex, non-linear relationships, offering highly detailed predictions for temperature, wind speed, and precipitation.

Key Points:

- Captures non-linear relationships and subtle patterns.
- Common algorithms: SVM**,** Random Forests**,** Neural Networks**.**
- Ideal for nowcasting and predicting extreme events.

1.3.3. Statistical Methods

Weather forecasting using statistical approaches focus on studying historical data to detect correlations and trends, often combining these methods with other algorithms to enhance accuracy [29]. For Example:

- **Regression Analysis:** Studying connections between factors, like the impact of humidity on the temperature fluctuations.
- **Ensemble Methods:** Integrates results from various models to minimize the errors and improve the reliability.

Key Points:

- Focus on historical data and trends.Concentrates on historical records and trends.
- Examples: Regression Analysis**,** Ensemble Methods**.**
- Enhances forecasting based on seasonal patterns and trends.

1.3.4. Kalman Filtering

Kalman filtering is a technique for adjusting weather forecasts by continuously integrating real-time observational data. When new data is received, the filter corrects predictions to address any differences between the model's predictions and observed observations, ensuring accurate weather conditions [30].

Kalman filtering is highly efficient in numerical weather prediction models, where continuous updates are vital in making constant adjustments for greater accuracy [31], [32].

Key Points:

- Constantly revises predictions based on real-time data inputs.
- Assures accuracy by minimizing variations and inconsistencies.
- Fundamental for optimizing NWP models.

1.4. Benefits of Weather Forecasting:

Disaster Management Preparation: Weather predictions support disaster readiness by delivering advance warnings for events such as hurricanes, floods, and tornadoes. They enable quick evacuations, effective resource distribution, and protection strategies, saving lives, limiting damage, and enhancing recovery efforts.

Economic Benefits: Weather forecasting provides valuable insights for agriculture, energy, and transportation, by supporting better crop management, optimizing power demand and renewable production, and minimizing delays and costs in transportation. The outcome is increased productivity, lower risks, and more efficient resource allocation, contributing to economic growth.

Safety: Weather forecasting enhances safety in aviation, maritime, and road transport by predicting dangers like turbulence, storms, and icy conditions. It helps in route optimization lower accident rates, and ensure secure travel for people and goods.

Improved Decision-Making: Weather forecasting provides crucial insights to help governments, businesses, and individuals make informed decisions. It supports construction, event planning, urban development, and personal choices like travel and clothing. This boosts decision-making, enhances resilience, and improves operational efficiency.

1.5. Limitations of Weather Forecasting:

Data Gaps: A major challenge that is faced in weather forecasting is data gaps in remote locations like oceans and mountains, where observational infrastructure is limited; the accuracy of forecasts is reduced, particularly for events like tropical cyclones. To overcome these challenges, additional infrastructure like buoys and drones is needed, which can be expensive.

Complexity: The unpredictable behavior of the atmosphere presents challenges for weather forecasting, as small changes can create significant variations in predictions. This affects the accuracy, particularly for localized events like storms, and even advanced forecasting models are influenced by oversimplifications and errors.

Cost: The infrastructure needed for weather forecasting is costly, requiring advanced technology such as high-performance computing, weather stations, radars, and satellites. Developing nations may face financial challenges, leading to disparities in access. Additionally, the costs of meteorologist training and system maintenance further hinder the global spread of advanced forecasting systems.

Accuracy Issues: Weather forecasting accuracy drops for long-term forecasts and extreme events due to the intricate behavior of the atmosphere and model limitations. This can result

in either under preparation or unnecessary precautions, leading to financial damage and weakening public confidence.

2. LITERATURE REVIEW

The literature review was conducted by analyzing a wide range of recent research papers related to weather prediction and machine learning applications. These studies include both region-specific investigations—some focusing on international contexts like Africa, Australia, United Kingdom, and the United States—and others rooted in diverse environmental conditions. While the global insights provide valuable methodological and technological perspectives, the primary focus of our study remains on the Indian climate. Emphasis has been placed on tailoring prediction models to India's unique meteorological patterns and challenges to ensure higher relevance and applicability of the findings.

Khatri et al. (2025) proposed a machine learning-based approach to enhance the accuracy of weather forecasting, particularly focusing on rainfall and flood prediction in urban regions. The study utilized historical meteorological data from 2009 to 2023, collected from various **Indian cities including Mumbai, Jaipur, Pune, Bengaluru, and Hyderabad**. Supervised learning models, including regression and ensemble methods, were employed after rigorous data preprocessing involving handling of missing values, outlier detection, and normalization. Python libraries such as Matplotlib, Pandas, and NumPy were used for data manipulation and visualization. The dataset comprised approximately 96,000 hourly temperature records with 25 attributes, covering a wide geographical range. The models were evaluated using performance metrics such as Accuracy, MAE, and MSE. Among the tested models, the LSTM model achieved the highest accuracy of 80.11%, particularly excelling in short-term rainfall prediction for flood-prone urban areas like Mumbai. The study demonstrated that machine learning models, particularly LSTM, can deliver high forecasting accuracy while remaining computationally efficient. These findings support the use of scalable, cost-effective machine learning frameworks for localized weather prediction, with potential applications in urban planning and disaster management [33].

Devapogu et al. (2025) proposed a comprehensive machine learning (ML) and deep learning (DL) framework for forecasting key weather parameters such as temperature, humidity, and wind speed. The study utilized a **weather dataset from Unagatla village, Andhra Pradesh, India,** covering daily records from January 2006 to February 2023. Preprocessing steps involved converting date information into temporal features like month, season, and day to capture seasonal trends. Exploratory data analysis, including trend and correlation studies, guided model development. Various ML models such as Random Forest Regressor, SVM,

XGBoost, and KNN were employed and evaluated using performance metrics. Although traditional models showed promising results especially Random Forest for maximum temperature they were less effective in forecasting humidity and precipitation. To address these limitations, advanced DL models like LSTM and hybrid architectures including CNN+LSTM and CNN+LSTM with Attention were implemented. These hybrid models demonstrated superior accuracy and generalization, particularly in predicting maximum and minimum temperatures and wind speed. The study concluded that hybrid DL models outperformed traditional ML approaches, offering improved reliability and robustness in weather forecasting applications [34].

Ramana et al. (2025) conducted a comprehensive investigation into the use of advanced deep learning (DL) and traditional machine learning (ML) techniques for multivariate weather forecasting using time series data. The study utilized a rich **dataset from Unagatla village, Andhra Pradesh, India,** comprising daily weather records from 2006 to 2022, including parameters such as maximum and minimum temperature, humidity, wind speed, and precipitation. Initially, traditional ML regression models like RFR, DTR, GBR, and MLR were employed as baseline models. Subsequently, deep learning models such as RNN, LSTM, and GRU networks were applied. The evaluation was based on performance metrics including MSE, MAE, RMSE, and R-squared (R^2) values. Results indicated that deep learning models, particularly LSTM, significantly outperformed traditional ML models across multiple weather parameters. LSTM achieved the highest accuracy in forecasting maximum and minimum temperatures, with MAE values of 0.032 and 0.031 and R^2 scores of 0.89 and 0.93, respectively. It also performed exceptionally well in predicting humidity, attaining an R^2 of 0.94 and MAE of 0.028, showcasing its effectiveness in capturing long-term dependencies in weather patterns. GRU models demonstrated superior performance in precipitation prediction with an MAE of 0.027 and R^2 of 0.95, balancing computational efficiency with strong predictive power. Overall, the study highlights the advantages of deep learning methods, especially LSTM and GRU, in enhancing the accuracy of multivariate weather forecasting [35].

Kaushal et al. (2025) conducted a detailed investigation into the combined effects of anisotropic foliations and weathering on the strength and deformation behavior of Himalayan gneiss rocks from **Kullu, Himachal Pradesh, India**. The study focused on three grades of weathered gneiss fresh, slightly weathered, and moderately weathered and examined their mechanical responses through extensive experimental testing, including uniaxial and triaxial

compression, Brazilian tensile, and point load strength tests. These tests were conducted across five anisotropy angles (0°, 30°, 45°, 60°, and 90°), revealing that both anisotropy and weathering significantly reduce rock strength under confined and unconfined conditions. A consistent U-shaped strength anisotropy curve was observed across all weathering grades, highlighting the influence of foliation angle on mechanical behavior. Importantly, the study introduced a novel correlation between triaxial strength and chemical weathering indices PIA, CIW and LOI derived from powder XRF analysis. These correlations allow for strength estimation at various anisotropy angles using chemically derived data, particularly useful when intact core specimens are unavailable. Furthermore, the study proposed a modified Hoek-Brown failure criterion that incorporates the effects of weathering via these chemical indices, achieving a high coefficient of determination ($R^2 = 0.93$). This approach presents a practical and innovative method for assessing anisotropic rock strength in the field of rock engineering, especially in challenging Himalayan terrains where sample preparation and availability can be constrained [36].

Kumar et al. (2024) highlighted the significance of weather prediction across sectors such as agriculture, transportation, and disaster management, emphasizing the need for accurate forecasts to support proactive decision-making. The study explored the use of ridge regression for predicting the next day's maximum temperature using historical weather data **from Bangalore, India**. Unlike traditional numerical models that are computationally intensive, the ridge regression model offers a more efficient alternative by effectively managing issues like multicollinearity and overfitting. The research placed strong emphasis on feature engineering, incorporating techniques such as rolling window statistics and expanding averages to improve predictive performance. Evaluation of the model was carried out using metrics like MAE, MSE and R-squared (R^2). The results demonstrated that ridge regression provided accurate and reliable temperature forecasts, effectively capturing daily temperature variations. The study also underscored the advantages of ridge regression, including its regularization capability, robustness against noise, and ability to maintain interpretability while managing the bias-variance trade-off. Overall, the work validates the applicability of machine learning, particularly ridge regression, in developing efficient and scalable weather prediction models for localized use [37].

Kaviyaraj et al. (2024) investigated the application of machine learning (ML) techniques to improve the accuracy of weather forecasting in coastal cities like Chennai, which are highly vulnerable to extreme weather events such as monsoons, cyclones, and urban flooding. The

study utilized an extensive **dataset comprising over 40,000 records** from climate measuring stations in **Chennai**, containing parameters such as rainfall, temperature, and geo-location data. Various ML models, including deep learning architectures like feed-forward neural networks, were applied to predict weather parameters such as temperature, precipitation, wind speed, and humidity. The performance of these models was compared with traditional numerical weather prediction techniques. The results demonstrated that ML models achieved higher precision and better adaptability to local climatic variations, particularly in short- to medium-term forecasts. Furthermore, the study highlighted the potential of integrating ML models with real-time data streams to develop adaptive, self-learning forecasting systems. While the models performed well overall, limitations were noted in capturing extreme weather conditions, suggesting the need for incorporating additional atmospheric variables such as air pressure and wind direction. The research also emphasized the future potential of hybrid models that combine physical weather equations with AI techniques to enhance forecasting accuracy. This study underscores the value of machine learning, especially deep learning, in enhancing disaster preparedness and urban planning strategies in climate-sensitive regions like **Chennai** [38].

Sharma et al. (2024) presented an innovative approach to weather forecasting by integrating machine learning algorithms such as Naive Bayes and Chi-Square, emphasizing their potential to overcome the limitations of traditional numerical weather prediction models. These traditional models often struggle with measurement uncertainties and the inherent complexity of atmospheric systems. The study utilized real-time, user-provided transformed data to eliminate reliance on historical datasets, demonstrating the feasibility of live forecasting using data mining techniques relevant to **Indian climatic conditions**. Among various models tested, including linear regression, multiple linear regression, decision tree regression, and polynomial regression, Random Forest Regression consistently outperformed others due to its high adaptability and precision across diverse weather scenarios. However, the study also revealed that linear and practical regression models were either too sensitive to outliers or unable to improve with additional data, highlighting the need for careful model selection and sufficient historical data length. Furthermore, the researchers suggested that extending the prediction window from two to five days could potentially reduce model bias, though it would require retraining and increased computation. Overall, the research underscores the transformative potential of combining ML techniques with traditional models to enhance weather prediction accuracy, especially for operationally critical sectors such as

agriculture, transportation, and disaster management. The findings support the continued optimization and real-time integration of machine learning methods to deliver more reliable and actionable weather forecasts [39].

Gahwera et al. (2024) emphasized the growing importance of short-term rainfall prediction in the context of climate change, highlighting the limitations of traditional numerical weather prediction models due to their high computational demands and susceptibility to errors. The study introduced a specialized Lake Victoria Basin weather dataset and investigated the efficacy of six machine learning regression models: Random Forest Regression, SVR, NNR, LASSO Regression, GBR and XGBoost in predicting hourly precipitation. The analysis of key environmental features, such as temperature, humidity, and wind direction, using the Pearson correlation coefficient, revealed weak linear relationships with precipitation. This finding necessitated the application of non-linear machine learning models to achieve more accurate predictions. Among the evaluated models, XGBoost demonstrated superior performance, achieving the lowest MAE values of 0.006, 0.018, and 0.005 for data from Uganda, Kenya, and Tanzania, respectively. The study underscores the potential of advanced machine learning techniques in enhancing rainfall predictions, which is critical for sectors such as agriculture, flood management, and aviation. By addressing the challenges posed by high rainfall intensity, this research offers valuable insights for improving environmental resilience and disaster preparedness [40].

Ashfaq et al. (2024) highlighted the critical need for accurate and timely wheat yield prediction, given the challenges posed by population growth, climate change, and increasing food demand. Their study focused on combining climate, soil, socioeconomic, and satellite data using three ML models SVM, RF and LASSO to predict wheat yield in the Multan region of Pakistan. Data from 2017 to 2022 was analyzed within the Google Earth Engine (GEE) platform to develop empirical models. The study found that integrating multiple data sources improved yield prediction significantly compared to using climate or satellite data alone. Among the ML models, RF outperformed SVM and LASSO, achieving the highest accuracy (R^2 = 0.88) and the lowest RMSE (0.05 q/ha). Combining spatial and weather-related predictors further enhanced accuracy, with RF achieving 97% accuracy in yield estimation. Non-linear models, particularly those using water-related predictors, performed better (R^2 = 0.82–0.92) than temperature-related predictors (R^2 = 0.34–0.63). The findings demonstrate the effectiveness of ML approaches, especially RF, in improving wheat yield

prediction accuracy, providing a robust alternative to traditional methods for the Multan district of Pakistan [41].

Kim et al. (2024) developed a machine learning model using five years of data from Seoul Grand Park to predict visitor numbers based on weather and day characteristics. Using the AdaBoost model, the study showed that incorporating weather variables like wind speed, humidity, and fine dust levels improved prediction accuracy. The model helps optimize theme park operations by enhancing resource allocation, marketing strategies, risk management, and financial forecasting, ultimately improving customer satisfaction and operational efficiency [42].

Bastian et al. (2024) developed a framework for improving weather forecasting at US military installations using high-resolution weather models. The Virtual Toolbox for Installation Mission Effectiveness (VTIME) combines real-time weather data with other installation information to assist decision-making. They used the Weather Research and Forecasting (WRF) model at a 1 km resolution to create more accurate forecasts, focusing on Fort Moore, GA. The study highlights the potential benefits of high-fidelity weather predictions but recommends further verification, automation, and integration with existing tools to enhance operational effectiveness at military sites [43].

Bouallègue et al. (2024) highlighted the growing potential of ML in weather forecasting, offering a cost-effective alternative to traditional numerical weather prediction. ML models, trained on high-quality reanalysis datasets like ERA5, provide accurate forecasts with significantly lower computational requirements compared to NWP systems. The study compared ML-based forecasts (e.g., PanguWeather) with traditional NWP forecasts (e.g., ECMWF IFS) using the same initial conditions in an operational context. Results showed comparable accuracy in global metrics and extreme event prediction, though ML models exhibited limitations such as smooth forecasts, increasing bias with lead time, and challenges in predicting tropical cyclone intensity. These findings indicate that ML could revolutionize weather forecasting by complementing or enhancing existing NWP methods [44].

Patil et al. (2024) emphasized the transformative role of machine learning (ML) across various domains, including weather forecasting. ML techniques such as Deep Learning, Random Forest, ANN and SVM have shown promise in predicting meteorological parameters like wind speed, pressure, precipitation, and temperature. However, due to the complexity and diversity of meteorological data, no single ML algorithm is universally effective for all

parameters. The study highlighted the advantages of hybrid ML techniques, ensemble models, and deep learning strategies, which combine outputs from multiple models to improve prediction accuracy. The effectiveness of ML-based weather forecasting is influenced by historical data, the time frame, and the location of weather stations. This flexibility and adaptability of ML underscore its importance as a key tool for improving the accuracy and reliability of weather predictions [45].

Arakh et al. (2024) explored the integration of Machine Learning (ML) algorithms such as Decision Trees, Random Forest, SVM, Neural Networks, and Gradient Boosting in weather forecasting. They highlighted the ability of ML to address limitations of traditional methods by learning complex data patterns, resulting in improved accuracy, efficiency, and scalability. The study emphasized the importance of feature selection, data preprocessing, and model evaluation, supported by real-world case studies demonstrating ML's advantages over conventional techniques. The paper also identified knowledge gaps and proposed future research directions, focusing on enhancing weather forecasting accuracy and societal preparedness for weather events. It showcased ML's transformative potential to deliver more accurate and timely predictions, benefiting decision-making and readiness in response to weather changes [46].

Wang et al. (2024) addressed the impact of convective weather on flight delays in terminal airspace and the need for accurate short-term forecasting to improve operational efficiency. They proposed a CNN-Transformer-based model using supervised deep learning, integrating eight weather parameters from civil aviation radar and ERA5 datasets. Trained on 241 datasets (23,881 samples), the model was validated through ablation studies and compared with existing forecasting methods. Results showed that while traditional radar extrapolation performed better for forecasts within one hour, the CNN-Transformer model excelled in 2-6 hour predictions, enhancing short-term convective weather forecasting for civil aviation [47].

Xu et al. (2024) investigated short-term precipitation forecasting in northern Xinjiang, where increasing rainfall heightens disaster risks due to fragile ecology and low infiltration capacity. They developed the PFformer model, a deep learning-based approach that reduces algorithm complexity using piecewise attention and enhances self-attention mechanisms. Utilizing data from five meteorological stations, the model was evaluated against mainstream time-series prediction models using MSE and MAE. Results showed PFformer improved MSE by 7.85%

and MAE by 9.74% on average, demonstrating its potential for enhancing precipitation forecasting accuracy and aiding disaster mitigation efforts [48].

Bojesomo et al. (2024) explored the use of artificial intelligence for short-term weather forecasting by leveraging a video transformer network. To overcome the computational complexity and data requirements of vision transformer architectures, they proposed a Video Swin-Transformer (VST) with a specialized augmentation scheme, gradual spatial reduction in the encoder, and cross-attention in the decoder. The model was tested on the Weather4Cast2021 dataset, predicting future frames 8 hours ahead. Results showed an MSE score of 0.4750 with training data and 0.4420 in a transfer learning scenario, demonstrating its potential for improved weather forecasting [49].

Xu et al. (2024) addressed challenges in multiple weather condition prediction by proposing a tensor-based long- and short-range convolution (TLS-Conv) operator. This approach integrates node & relation attention to identify spatial grid point contributions and utilizes adaptive tensor graph convolution (ATGCN) to capture long-range dependencies dynamically. By combining ATGCN with traditional convolution, the TLS-Conv-based TLS-MWP model effectively models spatial dependencies and weather correlations. Experimental results demonstrated that TLS-MWP outperforms existing methods, highlighting its potential for improving multi-condition weather forecasting [50].

Kumari et al. (2024) explored the use of the SARIMA model for weather forecasting, highlighting its effectiveness in predicting both short-term and long-term weather trends. SARIMA, an extension of the ARIMA model, incorporates seasonal components to account for periodic variations in weather data. The study demonstrated the model's high accuracy, achieving a RMSE of 1.24. By training on historical weather data, SARIMA proved to be a reliable machine learning approach for forecasting weather patterns, with its performance dependent on the quality of input data and model adjustments [51].

Ciszynski et al. (2024) analyzed the application of artificial intelligence in short-term weather forecasting using the ERA5 dataset from the European Centre for Medium-Range Weather Forecasts. The study focused on 24-hour forecasts for a specific location, optimizing models through hyperparameter tuning. Various machine learning models, including Support Vector Regression (SVR), logistic regression, stochastic gradient descent (SGD), K-nearest neighbors, Gaussian regression, decision tree, and random forest, were evaluated for their predictive performance [52].

Pantola et al. (2024) examined weather forecasting for Helsinki, a city known for its significant seasonal variations, using time series analysis. The study compared ARIMA, SARIMAX, LSTM, and GARCH models based on data from 2015 to 2019. Evaluating performance using MAE, MSE, RMSE, R^2 score, and MAPE, the results indicated that SARIMAX outperformed other models, making it the preferred choice for capturing Helsinki's temporal and geographical weather intricacies. The study highlights the importance of high-resolution regional forecasting models for applications in agriculture and emergency preparedness, offering more precise short- to medium-range weather predictions [53].

Sharma et al. (2023) addressed the critical challenge of real-time prediction of Heavy Rainfall Events (HREs) in complex terrains such as **Assam, India**, which are prone to flash floods and associated disasters. Motivated by the devastating HREs in June 2022 and June 2023 that caused extensive damage, displacement of millions, and significant economic losses, the study explored the use of deep learning (DL) techniques specifically U-Net architectures to enhance rainfall forecasting accuracy. Using simulated daily rainfall data from various parameterization schemes, the proposed U-Net model was evaluated over four days in June 2022. It outperformed traditional models like WRF, achieving a MAE below 12 mm and reducing error by 64.78% across **Assam**. The U-Net model was especially effective in capturing spatial rainfall variability and accurately predicting HREs in **districts like Barpeta, Kamrup, Kokrajhar, and Nalbari**. Moreover, in categorical rainfall prediction, the DL model achieved a high accuracy of 77.9%, outperforming the WRF models by 38.1%. The model's predictions closely aligned with observations from the India Meteorological Department (IMD), effectively capturing rainfall intensities above 150 mm, while the WRF models underestimated these events. The study concluded that U-Net models, particularly those with residual connections, provided superior performance in district-level rainfall forecasting, offering enhanced accuracy in both spatial and categorical dimensions. The research highlighted the potential of DL models in improving disaster preparedness and flood mitigation strategies. Future work aims to further enhance model robustness by training on larger datasets to better capture complex rainfall patterns and improve prediction reliability [54].

Choubisa et al. (2023) addressed the long-standing challenge of accurate weather forecasting, highlighting that despite decades of advancements, traditional forecasting methods still fall short in precision. Given the critical dependence of agricultural and

industrial sectors on weather conditions, reliable forecasting has become increasingly essential, not only for daily operations but also for disaster prediction and mitigation. Weather forecasting involves estimating future atmospheric conditions based on a range of meteorological parameters. In their study, Choubisa et al. applied deep learning techniques, specifically LR and DNN to enhance forecasting accuracy for the **Jaipur district in Rajasthan, India**. The **dataset, sourced from Kaggle,** spans two years (May 1, 2016, to March 11, 2018) and includes 40 features with 676 instances, covering parameters like temperature, humidity, dew point, and precipitation. Due to the highly dynamic and irregular nature of weather patterns, accurate forecasting remains a complex task. The study underscores the importance of using robust models and high-quality datasets to overcome these challenges. Their experimental results demonstrate that while both LR and DNN can efficiently handle high-dimensional meteorological data, DNN outperforms LR in terms of prediction accuracy and learning efficiency. However, the limited dataset size also points to the need for further research and larger datasets to enhance model generalizability [55].

Suthar et al. (2023) investigated the increasing severity and frequency of heat waves, primarily driven by global warming, with a specific focus on the **Indian states of Rajasthan and Karnataka**. These regions have been experiencing recurrent extreme heat events in recent years. The study utilized machine learning techniques to predict the maximum air temperature (AT), a critical parameter for identifying heat wave occurrences. Using monthly data of 13 environmental parameters collected over the summer seasons from 2013 to 2022, **sourced from NASA's Giovanni and ERA5 reanalysis datasets**, the data was resampled to a spatial resolution of 5 km × 5 km. Pearson's correlation and sensitivity analysis helped identify dominant parameters affecting AT, including land surface temperature, black carbon, aerosol optical depth and carbon monoxide with LST and BC having the highest influence. The study employed three machine learning models MLR, SVR, and RF and validated their performance using 3-fold cross-validation. The Random Forest model outperformed the others, achieving high predictive accuracy with adjusted R^2 values of 0.90 for Rajasthan and 0.92 for Karnataka, and RMSE values of 2.36% and 1.44%, respectively. These findings demonstrate the effectiveness of machine learning in predicting maximum AT and assessing heat wave risk. The study holds valuable implications for climate modeling, urban infrastructure planning, and agricultural management, offering a data-driven basis for formulating adaptive strategies to mitigate the impacts of rising temperatures [56].

Verma et al. (2023) emphasized the critical need for accurate thunderstorm prediction due to the severe and hazardous weather conditions such phenomena can cause. Their study focused on forecasting **rainfall and thunderstorms in Jaipur, Rajasthan**, using various machine learning techniques. Among the models evaluated, Decision Trees, SVM, KNN, AdaBoost, and Gradient Boosting the Random Forest Classifier consistently outperformed the others, achieving remarkable accuracy levels ranging from 92.55% to 98.92% for predicting events such as "Rain," "Thunderstorm," and "Thunderstorm with Rain." The robustness and predictive accuracy of the Random Forest model position it as an effective tool for weather forecasting and climatic analysis, with implications across sectors that depend heavily on precise meteorological data. The study utilized a rich dataset spanning 48 years (1969–2017), **sourced from the India Meteorological Department (IMD)**, which included daily summaries of weather parameters such as temperature, rainfall, wind speed, evaporation, sunshine hours, and various weather events like fog, hail, and drizzle. The results underline the potential of machine learning, especially ensemble methods like Random Forest, to enhance the accuracy and reliability of long-term weather predictions. This research significantly contributes to climatology and meteorology by improving the understanding of weather trends in Jaipur and laying the groundwork for further advancements in data-driven weather forecasting [57].

Hassan et al. (2023) This study applied machine learning algorithms for accurate rainfall prediction, using techniques like data preprocessing, feature selection, and models including Naive Bayes, SVM, Random Forest, and ANN. ANN achieved the highest accuracy of 91%. Regional rainfall patterns in Australia were analyzed using k-means clustering and PCA, and a Flask-based web application was developed for public access. The research highlighted the potential for hybrid classifiers and advanced techniques to improve prediction accuracy. Future directions include using larger datasets, exploring deep learning models, implementing federated learning, and addressing limitations such as reliance on historical weather data and regional data variability. This work underscores the promise of machine learning in enhancing rainfall forecasting systems [58].

Tanabe et al. (2023) evaluated the effectiveness of CNNs for crop yield prediction in small-scale Asian agricultural fields. Using UAV-based multispectral imagery captured during four growth stages of winter wheat, the study assessed how growth stages and their combinations impact prediction accuracy. The CNN model performed best during the heading stage, achieving the lowest RMSE (0.94 t/ha), outperforming linear regression models based on the

enhanced vegetation index 2 (EVI2). Multi-temporal CNN and linear regression models showed lower accuracy compared to single-stage CNN models. These findings highlight that the heading stage is optimal for data collection, and combining multiple growth stages does not enhance accuracy. The study suggests that further improvements could be achieved with higher-resolution imagery and integration of weather data. Overall, the CNN model proved to be a precise and efficient method for yield prediction, reducing the need for frequent UAV data acquisition [59].

Gangula et al. (2023) developed a machine learning-based weather prediction model using five algorithms. Among them, Random Forest achieved the highest accuracy of 75%. The research also highlights the challenges posed by the complex and chaotic nature of atmospheric systems in achieving precise forecasts. It also notes that broader weather classifications yield higher prediction accuracy, while increased specificity reduces precision. This work underscores the importance of balancing classification detail and accuracy in weather forecasting models [60].

Iram et al. (2023) emphasized the importance of predicting energy consumption to enhance energy conservation and reduce costs in smart homes. The study applied six machine learning algorithms to forecast energy use based on weather data. PCA was used for dimensionality reduction, and Lasso Regression identified key weather-related patterns affecting energy consumption. A decision matrix with fuzzy operators ranked the models, with SVR achieving the highest accuracy after resampling. The study highlights the potential of machine learning in optimizing energy usage by integrating weather information for smarter energy management [61].

Chen et al. (2023) reviewed the growing role of machine learning (ML) in meteorology, highlighting its potential for short-term weather and medium-to-long-term climate forecasting. The study analyzed over 20 methods, focusing on eight promising techniques that improve forecasting accuracy. While ML excels in short-term weather predictions, its application in long-term climate forecasting faces challenges due to complex climate variables and data limitations. The survey emphasized the need for integrated research connecting weather and climate forecasting and addressing gaps in modeling structure and data availability. The study provided a detailed evaluation of key models, serving as a guide for model selection in different contexts. It identified challenges like limited seasonal datasets and proposed future research directions, such as data simulation and incorporating physics-

based constraints. This work offers valuable insights into the transformative potential of ML in meteorology and outlines pathways for advancing interdisciplinary research in this field [62].

Mung et al. (2023) explored deep learning techniques to forecast minimum and maximum temperature, humidity, and wind speed using 20 years of daily weather data from Yangon, Myanmar. It compared CNN, LSTM, and a hybrid CNN-LSTM model for predictive performance. Performance evaluation using Root Mean Square Error (RMSE) demonstrated that the ensemble CNN-LSTM model outperformed individual CNN and LSTM models, providing more accurate weather predictions [63].

Zenkner et al. (2023) investigated the use of deep learning for numerical weather prediction by developing two Bi-LSTM recurrent neural network models trained on historical weather data from two locations in London. The first model predicted 24-hour temperature at Kew Gardens with 73% accuracy within ±2°C and a RMSE of 1.45°C. The second model forecasted 72-hour air temperature and relative humidity at Heathrow, achieving RMSEs of 2.26°C and 14%, respectively. Trained on five years of data, both models demonstrated promising performance, with cloud training times of approximately one and three minutes for the 24-hour and 72-hour networks, respectively [64].

Choi et al. (2023) introduced a novel deep learning-based climate model combining the NODE and the advection diffusion equation, widely used in climate modeling to describe physical processes like Brownian and bulk motions. Their proposed framework, called the Neural Advection Diffusion Equation (NADE), incorporates an additional neural network to address uncertainty and learn the best latent governing equation for climate data. Experiments using real-world and synthetic datasets demonstrated that NADE outperformed 14 baseline models, highlighting its effectiveness in climate modeling [65].

Najib et al. (2022) developed a local weather prediction system utilizing IoT devices and an ANN model to aid rice farmers in managing weather-related challenges. The ANN model, trained with normalized weather data, achieved optimal accuracy using 40 training cycles and seven hidden layers. The system was tested with September 2021 data for training and October 2021 data for validation. The model demonstrated high accuracy across various weather parameters, such as temperature (90.5%), humidity (87.2%), rainfall (68.2%), wind speed (89.6%), and light intensity (83.1%). The study highlighted the importance of normalization to balance input data and ensure accurate predictions. The system successfully

forecasted weather conditions for October 1-3, 2021, showing temperature increases, reduced humidity, and variable wind speed and light intensity. This approach demonstrates the potential of combining IoT and ML for localized, accurate weather predictions to support agricultural decision-making [66].

Fowdur et al. (2022) proposed a collaborative ML-based system was developed for real-time, localized weather forecasting in Mauritius using data from multiple locations. Five algorithms MLR, MPR, KNN, MLP, and CNN were used to predict temperature, wind speed, pressure, and humidity. Data was sourced via the OpenWeather API and analyzed on local and cloud servers. Results showed collaborative regression schemes achieved 5% lower MAPE compared to non-collaborative methods, with MPR yielding the highest accuracy (errors ranging from 0.009% to 9%). This highlights the potential of collaborative approaches to improve weather forecasting accuracy [67].

Aderyani et al. (2022) short-term rainfall forecasting was evaluated using PSO-optimized SVR, LSTM, and CNN with data from Niavaran station, Tehran. PSO-SVR performed best for 15-minute predictions, while LSTM excelled in 5-minute forecasts, outperforming CNN overall.. Event classification based on severity and duration improved accuracy, while incorporating additional predictors, such as rainfall depth fluctuations, further enhanced the PSO-SVR and LSTM models' performance by 3–15% and 2–10%, respectively [68].

Meenal et al. (2022) highlighted the importance of accurate weather forecasting for sectors like agriculture and aviation. It predicts parameters such as precipitation, humidity, wind speed, and solar radiation across **Indian locations** using traditional temperature-based models (e.g., Hargreaves-Samani, Bristow-Campbell) and ML algorithms like linear regression, SVM, and decision tree. Experimental **data from the Indian Meteorological Department (IMD)** validated the models. Results showed that ML models outperformed empirical methods with a correlation coefficient of 0.9259 and a mean square error (MSE) of 0.1397, providing better predictions with lower computational effort. This work demonstrated the potential of ML in optimizing weather forecasts, particularly for agriculture, where accurate predictions can improve crop yield and planning [69].

Bochenek et al. (2022) analyzed 500 key research articles published since 2018 on machine learning applications in climate and numerical weather prediction. The study highlighted common topics such as renewable energy forecasting, atmospheric physics, climate change, and extreme weather events. Key meteorological fields like wind, precipitation, and

temperature were frequently studied using methods such as Deep Learning, Random Forest, Support Vector Machines, and XGBoost. The authors emphasized that understanding relationships between meteorological variables and proper implementation of machine learning techniques are crucial for effective forecasting. They concluded that machine learning will play a significant role in future weather prediction, aiding in bias correction, data assimilation, and high-resolution modeling. Additionally, these methods hold potential for climate change research, depending on the availability of high-quality data [70].

Gebretinsae et al. (2022) addressed the challenge of weather prediction for agriculture in Ethiopia, proposing ML models as an alternative to traditional methods. The study focused on five key weather parameters: Temperature, Precipitation, Sunshine Hours, Relative Humidity, and Rainfall, using data from 44 meteorology stations in the Adama region. The ML models, based on learning patterns from historical data, showed improved accuracy and simplicity compared to existing models, encouraging meteorology personnel to adopt these techniques. The research identified the limitations of traditional NWP models and demonstrated that the proposed ML models performed comparably to the existing WRF model. Time series models like EMA and ARIMA were found to be more effective than the moving average model, leading to the development of more efficient weather prediction tools for agricultural applications in Ethiopia. The results and scripts were shared with the Ethiopian Meteorology Centre (EMC) for integration into their weather forecasting tasks and agro-advisory development [71].

Faraji et al. (2020) A hybrid energy management system (EMS) was proposed for prosumer microgrids, using ANFIS, MLP-ANN, and RBF-ANN to forecast load and weather parameters. By selecting the best-performing model for each variable, the system achieved improved prediction accuracy and optimized microgrid operation. The proposed EMS incorporated a demand response (DR) program and considered battery storage system degradation costs. Results showed a 4.2% reduction in PMG operation costs using the DR program and hybrid forecasting method, achieving a cost of $1.07, compared to $1.349 with conventional methods. Additionally, power purchased from the main grid decreased by 8.55%, reducing emissions and enhancing cost efficiency. The research highlighted the effectiveness of hybrid forecasting methods and DR programs in optimizing PMG operations [72].

Hossain et al. (2020) developed an LSTM-based model for photovoltaic (PV) power forecasting using synthetic irradiance data generated from historical patterns and sky forecasts. K-means clustering improved accuracy by dynamically classifying sky conditions. The model achieved up to 44.6% higher accuracy than daily forecasts and outperformed RNN, GRNN, and ELM, highlighting LSTM's effectiveness for PV prediction [73].

Zhang et al. (2019) proposed a multimodal fusion method for predicting visibility, which combines numerical weather predictions with factory emission data from satellite images. Using advanced algorithms like XGBoost and LightGBM, the model's accuracy was improved compared to traditional methods. By incorporating emission data from Landsat-8 satellite images, the prediction system was further enhanced. The study suggests that adding more factors, beyond factory emissions, could further improve predictions. However, the need for high-resolution satellite images poses a challenge, and future work could explore additional data sources [74].

Singh et al. (2019) compared SVM, ANN, and Time Series RNN for weather prediction, evaluating performance using RMSE. The Time Series RNN achieved the best results with the lowest RMSE of 1.41 over an 8-week forecast period, outperforming SVM and ANN. The research highlights the importance of prediction window size, noting that larger windows lead to higher errors, and demonstrates the effectiveness of RNNs for accurate weather forecasting [75].

Aznarte et al. (2017) conducted a dynamic line-rating (DLR) study to predict ampacity for two conductor lines in Northern Ireland up to 27 hours in advance. The study utilized four machine learning algorithms generalized linear models, MARS, random forests, and quantile random forests combined with numerical weather predictions. MARS emerged as the most effective regression method, offering significantly improved point and probabilistic forecasts compared to standard models. The research highlighted the benefits of probabilistic forecasts, including safety margin computations to prevent risks. The findings demonstrated the feasibility of DLR forecasts for real-time monitoring and protecting transmission and distribution systems, especially with high renewable energy integration. The study suggested further refinements, such as input variable optimization, dynamic safety margins, and a comprehensive line-rating forecast approach. These advancements could enhance the operational use of DLR for network protection, capacity gains, and control room applications [76].

Sagar et al. (2017) addressed the critical issue of predicting extreme rainfall events, which are often precursors to severe flooding in **various regions, including India**. The study evaluated the performance of numerical weather prediction (NWP) models in forecasting rainstorms during the **Indian monsoon season (June to September)** using 15 identified rainstorm events based on gridded daily rainfall datasets and objective criteria. The analysis utilized three global ensemble models from the TIGGE (THORPEX Interactive Grand Global Ensemble) framework ECMWF, NCEP, and UKMO—covering 43 rainstorm days between 2007 and 2015. Results indicated that while these events are predictable up to five days in advance, significant biases in spatial distribution and intensity persist. Among the models, ECMWF and UKMO exhibited wider ensemble spreads compared to NCEP, though NCEP showed poorer performance in predicting ensemble averages. Statistical validation using mean error (Bias), RMSE and correlation coefficient revealed underprediction tendencies, as confirmed by rank histograms. The study further applied the modified CRA technique to assess both spatial and quantitative accuracy, highlighting that displacement and pattern errors significantly contributed to the total RMSE. Additionally, volume errors increased from 24-hour to 48-hour forecasts across all models. Overall, the findings emphasize the need for improvements in ensemble-based NWP systems to enhance the accuracy of extreme rainfall forecasting over the Indian subcontinent [77].

3. RESEARCH GAP

- **Long-Term Forecasting Accuracy:** Existing weather prediction models struggle with accurately forecasting long-term climate trends and extreme weather events, highlighting the need for advanced models that can adapt to evolving climate patterns.
- **Real-Time Model Adaptation:** Current machine learning and deep learning-based weather forecasting systems require improved real-time training and adaptation mechanisms to integrate newly available data dynamically and enhance predictive accuracy.
- **Extreme Weather Event Prediction:** There is a need for more robust forecasting models capable of handling the increasing frequency and intensity of extreme weather events, ensuring timely and reliable predictions for disaster preparedness and mitigation.

4. OBJECTIVE

1. To analyze historical weather data and identify patterns and trends using deep learning techniques.
2. To design and develop a weather prediction model using deep learning techniques.
3. To compare the performance parameters of existing forecasting models with the proposed model and optimize its performance

5. PROPOSED METHODOLOGY

To develop a Weather Prediction and Forecasting System using Machine Learning (ML) and Deep Learning (DL) techniques, we can follow these basic steps:

1. Define the Problem and Objectives

Weather prediction involves forecasting different weather parameters such as temperature, humidity, precipitation, wind speed, and atmospheric pressure. Defining the problem includes:

- **Identifying key weather variables**: Determine which parameters are crucial for prediction based on application (e.g., temperature for agriculture, wind speed for renewable energy).
- **Setting specific goals**:
  - Improve accuracy compared to traditional forecasting models.
  - Reduce computational costs for real-time predictions.
  - Enhance model interpretability and usability.
- **Determining forecast timeframes**: Short-term (hours to days), medium-term (weeks), or long-term (months to years).

2. Data Collection

The accuracy of weather predictions depends on the quality of the data collected. Data collection involves:

- **Gathering historical weather data** from reliable sources such as:
  - Government meteorological agencies (e.g., NOAA, IMD, ECMWF).
  - Satellite data providers (e.g., NASA Earthdata, MODIS, Copernicus).
  - IoT sensors and personal weather stations (e.g., OpenWeatherMap API, Weather Underground).
  - **Historical databases**:
    - Kaggle datasets
    - Global Surface Summary of the Day (GSOD)
    - ERA5 reanalysis dataset
    - open-meteo.com

- **Ensuring comprehensive data coverage**: Including temperature, humidity, wind speed, pressure, precipitation, solar radiation, and cloud cover.
- **Data cleaning and preprocessing**:
    - Handling missing values through interpolation or imputation techniques.
    - Removing outliers and inconsistent readings using statistical methods.
    - Converting categorical data (e.g., "sunny", "cloudy") into numerical values.

3. Data Preprocessing and Feature Engineering:

To ensure the data is structured optimally for modeling, preprocessing and feature engineering are required:

- **Data Normalization and Scaling**:
    - Min-Max Scaling: Suitable for deep learning models.
    - Standardization (Z-score normalization): Useful for tree-based models.
- **Feature Engineering**:
    - Creating time-related features (hour, day, month, season).
    - Generating lag features (past weather conditions as inputs).
    - Calculating rolling statistics (moving averages, variance, etc.).
    - Creating interaction features (e.g., temperature-humidity index).
- **Splitting the data** into:
    - Training set (70-80%): Used for model training.
    - Validation set (10-15%): Used for hyperparameter tuning.
    - Test set (10-15%): Used for final evaluation.

4. Exploratory Data Analysis (EDA):

EDA helps uncover patterns and trends in the data before model selection:

- **Data Visualization Techniques**:
    - Line charts for temperature trends.
    - Scatter plots to explore relationships (e.g., humidity vs. temperature).
    - Histograms to understand data distributions.
    - Box plots to detect outliers.

- **Correlation Analysis**:
    - Pearson correlation coefficient to measure dependencies.
    - Heatmaps to visualize relationships between multiple variables.
- **Time-Series Analysis**:
    - Seasonal decomposition to separate trend, seasonality, and residual components.
    - Autocorrelation plots to detect dependencies in time-series data.

5. **Select Models:** Choose Machine Learning (ML) models like Linear Regression, or Deep Learning (DL) models like LSTM, depending on the complexity of the task.
6. **Model Training and Validation:** Train the models on the data, ensuring generalization using cross-validation, and optimize hyperparameters for better performance.
7. **Model Evaluation:** Use metrics like MAE or RMSE to evaluate model accuracy and compare ML and DL results with traditional methods.
8. **Model Optimization:** Fine-tune models with regularization, dropout, or ensemble methods to improve performance.
9. **Implementation and Forecasting:** Integrate the models into a system that can predict weather based on real-time or historical data, ensuring user-friendliness and efficiency.
10. **Deployment and Monitoring:** Deploy the system for real-time predictions and continuously monitor and update models for improved performance.
11. **Visualization and Reporting:** Create visualizations and reports to display forecast results, trends, and insights for better decision-making.
12. **Model Evaluation and Iteration:** Continuously evaluate and refine the system by incorporating new data and feedback to improve accuracy and efficiency.

Flow Chart of Proposed Methodology

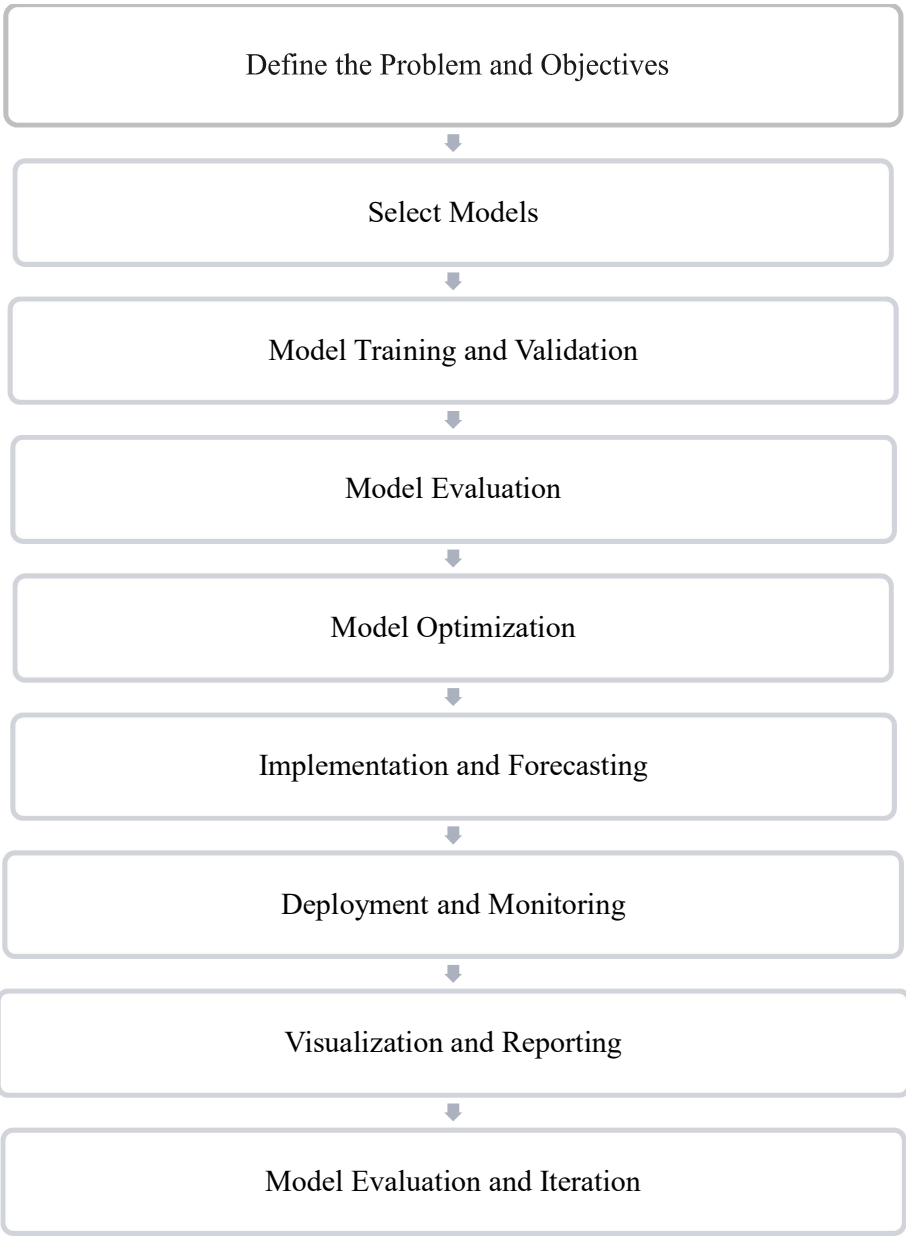

Figure 1 Flowchart of Proposed Methodology

6. WORK PLAN FOR THE RESEARCH

S. No.	Tasks	Plan of Work (in months)					
		0-6	7-12	13-18	19-24	25-30	31-36
1.	Course Work	Y	Y				
2.	Review of existing Work			Y			
3.	Simulation on Software or Model Design			Y	Y	Y	
4.	Performance analysis of the various model for weather prediction			Y	Y	Y	
5.	Submission of Proposal			Y			
6.	Communication of papers based on analysis outcomes			Y	Y	Y	Y
7.	Thesis Writing and Documentation					Y	Y

7. REFERENCES

[1]. O. A. Wani, S. S. Mahdi, Md. Yeasin, S. S. Kumar, A. S. Gagnon, F. Danish, N. Al-Ansari, S. El-Hendawy, M. A. Mattar, "Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas", Scientific Reports, Vol. 14, Article number: 27876, 2024.

[2]. A. Gupta, H. K. Mall and S. Janarthanan., "Rainfall Prediction Using Machine Learning," IEEE 2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR), pp. 1-5, 2022.

[3]. I. A. Kachalla, C. Ghiaus, M. Baseer, "Comparative analysis of machine learning models for prediction and forecasting of electric water boilers energy consumption", Applied Thermal Engineering, Vol. 267, pp. 1-22, 2025.

[4]. D. Markovics, M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction", Renewable and Sustainable Energy Reviews, Vol. 161, pp. 1-17, 2022.

[5]. H. Chen, Q. Zhang, Y. Birkelund, "Machine learning forecasts of Scandinavian numerical weather prediction wind model residuals with control theory for wind energy", Energy Reports, Vol. 8, pp. 661-668, 2022.

[6]. R. A. Schwalbert, T. Amado, G. Corassa, L. P. Pott, P. V. V. Prasad, I. A. Ciampitti, "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil", Agricultural and Forest Meteorology, Vol. 284, 2020.

[7]. M. M. Hassan et al., "Machine Learning-Based Rainfall Prediction: Unveiling Insights and Forecasting for Improved Preparedness", IEEE Access, vol. 11, pp. 132196-132222, 2023.

[8]. S. Ghaffarian, Fi. R. Taghikhah, H. R. Maier, "Explainable artificial intelligence in disaster risk management: Achievements and prospective futures", International Journal of Disaster Risk Reduction, Vol. 98, pp. 1-22, 2023.

[9]. O. Z. Romero, P. A. S. Chavez, P. C. Gonzalez, J. Z. Hidalgo, A. R. Garcia, P. O. Alfaro, R. R. Centeno, O. G. Ramos, "Operational ozone forecasting system in

Mexico City: A machine learning framework integrating forecasted weather and historical ozone data", Atmospheric Environment, Vol. 344, 2025.

[10]. Y. Tao, B. Tian, B. R. Adhikari, Q. Zuo, X. Luo, B. Di, "A Review of Cutting-Edge Sensor Technologies for Improved Flood Monitoring and Damage Assessment", Sensors, Vol. 24(21), pp. 1-27, 2024, doi: 10.3390/s24217090.

[11]. H. Chan, M.I.M. Wahab, "A machine learning framework for predicting weather impact on retail sales", Supply Chain Analytics, Vol. 5, pp. 1-14, 2024.

[12]. S. D. Latif, N. A. B. Hazrin, C. H. Koo, J. L. Ng, B. Chaplot, Y. F. Huang, A. El-Shafie, A. N. Ahmed, "Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches", Alexandria Engineering Journal, Vol. 82, pp. 16-25, 2023.

[13]. H. Qian, W. Wang, G. Chen, "Assessing forecast performance of daily reference evapotranspiration: A comparison of equations, machine and deep learning using weather forecasts", Journal of Hydrology, Vol. 644, 2024.

[14]. H-H Lin, C-C Tsai, J-C Liou, Y-C Chen, C-Y Lin, L-Y Lin, K-S Chung, "Multi-Weather Evaluation of Nowcasting Methods Including a New Empirical Blending Scheme", Atmosphere Special Issue Advances in Mesoscale Numerical Weather Prediction and its Applications, Vol. 11, 2020.

[15]. S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, R. Prudden, A. Mandhane, A. Clark, A. Brock, K. Simonyan, R. Hadsell, N. Robinson, E. Clancy, A. Arribas and S. Mohamed, "Skilful precipitation nowcasting using deep generative models of radar", Nature, Vol. 597, pp. 672-677, 2021.

[16]. G B Wanugroho, Martarizal, R M Putra, "Implementation of artificial neural networks for very short range weather prediction", Journal of Physics: Conference Series, Volume 1528, 4th International Seminar on Sensors, Instrumentation, Measurement and Metrology 14th November 2019, Padang, Indonesia, 2020.

[17]. J.Frnda, M. Durica, J. Rozhon, M. Vojtekova, J. Nedoma, R. Martinek, "ECMWF short-term prediction accuracy improvement by deep learning", Scientific Reports, 12, Article number: 7898, 2022.

[18]. L. Bengtsson, "Medium Range Forecasting at ECMWF; A Review and Comments on Recent Progress", Problems and Prospects in Long and Medium Range Weather Forecasting, Topics in Atmospheric and Oceanographic Sciences, pp. 21-41, 1984.

[19]. Kaifeng Bi, L. Xie, H. Zhang, X. Chen, X. Gu, Q. Tian, "Accurate medium-range global weather forecasting with 3D neural networks", Nature, Vol. 619, pp. 533-538, 2023.

[20]. M. Davey, A. Brookshaw, "Long-range meteorological forecasting and links to agricultural applications", Food Policy, Vol. 36, pp. 88-93, 2011.

[21]. H. Zhang, Y. Liu, C. Zhang, N. Li, "Machine Learning Methods for Weather Forecasting: A Survey", Atmosphere, Special Issue Applications of Artificial Intelligence in Atmospheric Sciences, 2025.

[22]. T N Palmer, "Predicting uncertainty in forecasts of weather and climate", Reports on Progress in Physics, Vol. 63, No. 2, 2000.

[23]. N. Logar, "Climate Prediction and Weather Forecasting", Applied Agrometeorology, pp. 959-964, 2010.

[24]. B. Goger, M. W. Rotach, A. Gohm, I. Stiperski and O. Fuhrer, "Current challenges for numerical weather prediction in complex terrain: Topography representation and parameterizations," 2016 International Conference on High Performance Computing & Simulation (HPCS), pp. 890-894, 2016.

[25]. S. K. Park, "Numerical Weather Prediction: East Asian Perspectives", Springer Atmospheric Sciences, eBook Packages Earth and Environmental Science, Earth and Environmental Science (R0), 2023.

[26]. A. Ayoub, H. M. Wainwright, G. Sansavini, "Machine learning-enabled weather forecasting for real-time radioactive transport and contamination prediction", Progress in Nuclear Energy, Vol. 173, 2024.

[27]. M. Waqas, U. W. Humphries, B. Chueasa, A. Wangwongchai, "Artificial intelligence and numerical weather prediction models: A technical survey", Natural Hazards Research, 2024.

[28]. C. Lalitha, D. Ravindran, "Hybrid deep learning framework for weather forecast with rainfall prediction using weather bigdata analytics", Multimedia Tools and Applications, Vol. 83, pp. 66413-66432, 2024.

[29]. D. Kampolis, "Statistical Analysis for Long-Term Weather Forecast", Proceedings of 16th International Conference on Meteorology, Climatology and Atmospheric Physics—COMECAP 2023, Environmental Sciences proceedings, 2023.

[30]. N. Bogdanovs, V. Bistrovs, E. Petersons, A. Ipatovs and R. Belinskis, "Weather Prediction Algorithm Based on Historical Data Using Kalman Filter," 2018 Advances in Wireless and Optical Communications (RTUWO), pp. 94-99, 2018.

[31]. F. Cassola, M. Burlando, "Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output", Applied Energy, Vol. 99, pp. 154-166, 2012.

[32]. B. Wang, Z. Sun, X. Jiang, J. Zeng, R. Liu, "Kalman Filter and Its Application in Data Assimilation", Atmosphere, Special Issue Data Assimilation Development: Theory, Algorithm, and Applications in Meteorology, 2023.

[33]. Sunil Khatri, Rajani P.K, "Advanced Weather Forecasting with Machine Learning: Leveraging Meteorological Data for Improved Predictions", Communications on Applied Nonlinear Analysis, Vol 32, No. 5s, pp. 1-16, 2025.

[34]. K. Devapogu, M. Venkatesh, N. Vullam, V. D. Babu, C. R. Naidu and A. Lakshmanarao, "A Novel Approach to Farm Weather Prediction with Hybrid CNN, LSTM, and Attention Mechanisms," 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), pp. 1-7, 2025.

[35]. C. V. Ramana, B. V. Varma, U. V. Ramesh, P. A. Kumar, A. Lakshmanarao and R. R. P. V, "Advanced Deep Learning Techniques for Multivariate Weather Prediction Using RNN and LSTM Models," 2025 Fifth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), pp. 1-6, 2025.

[36]. H. Kaushal, A. Singh, N. K. Samadhiya, "Evaluating the combined influence of anisotropy and weathering on strength behavior of gneissic rocks in Himachal Pradesh, India: Insights from chemical weathering indices", Engineering Geology, VOl. 352, 2025.

[37]. k. R. Suresh Kumar, P. Kambli, M. N. Prathima, G. Shruthi, S. Komal Raj and N. Poojari, "Weather Prediction Model using Ridge Regression," 2024 International Conference on Recent Advances in Science and Engineering Technology (ICRASET), pp. 1-5, 2024.

[38]. Kaviyaraj R, Karan S, John Paul R, Agilesh A, Madhesh H, "Deep Learning-Based Rainfall Prediction for Coastal Cities: A Case Study of Chennai," 2024 International BIT Conference (BITCON), pp. 1-6, 2024.

[39]. M. Sharma, R. P. Singh, U. Sharma, R. Arivukkodi, R. K. Roy and K. Verma, "Enhancing Weather Forecasting Accuracy: Integrating Machine Learning Algorithms with Numerical Weather Prediction Models," 2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N), pp. 1394-1399, 2024.

[40]. T. A. Gahwera, O. S. Eyobu and M. Isaac, "Analysis of Machine Learning Algorithms for Prediction of Short-Term Rainfall Amounts Using Uganda's Lake Victoria Basin Weather Dataset", IEEE Access, vol. 12, pp. 63361-63380, 2024.

[41]. M. Ashfaq, I. Khan, A. Alzahrani, M. U. Tariq, H. Khan and A. Ghani, "Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion," in IEEE Access, vol. 12, pp. 40947-40961, 2024.

[42]. J. Kim and S. Kim, "A Study on Estimating Theme Park Attendance Using the AdaBoost Algorithm Based on Weather Information from the Korea Meteorological Administration Web," in Journal of Web Engineering, vol. 23, no. 6, pp. 869-884, September 2024.

[43]. H. Y. Bastian, B. Hoch, I. V. Obiako, R. K. Buchanan, J. Richards and C. Rinaudo, "Framework for Numerical Weather Predictions for Smart Military Installations," 2024 IEEE International Systems Conference (SysCon), pp. 1-3, 2024.

[44]. Z. B. Bouallègue, M. C. A. Clare, L. Magnusson, E. Gascón, M. Maier-Gerber, M. Janoušek, M. Rodwell, F. Pinault, J. S. Dramsch, S. T. K. Lang, B. Raoult, F. Rabier, M. Chevallier, I. Sandu, P. Dueben, M. Chantry, F. Pappenberger, "The Rise of Data-Driven Weather Forecasting: A First Statistical Assessment of Machine Learning–Based Weather Forecasts in an Operational-Like Context", Bulletin of the American Meteorological Society, Vol. 105, Issue. 6, pp. 864-883, 2024.

[45]. Rohit P. Patil, KamalakarR.Desai, "Innovative Machine Learning Approaches for Prediction of Weather Parameters", Journal of Electrical System, Vol. 20, Isssue. 3, pp. 6820-6827, 2024.

[46]. R. Arakh, Y. Gupta, S. Kriplani, S. Sharma, "Advancements in Weather Forecasting through Machine Learning Algorithms", International Journal of Recent Advances in Multidisciplinary Topics, Vol. 5, Issue. 5, pp. 145-148, 2024.

[47]. S. Wang, Y. Li, B. Yang and R. Duan, "Short-Term Forecasting of Convective Weather Affecting Civil Aviation Operations Using Deep Learning," in IEEE Access, vol. 12, pp. 166011-166030, 2024.

[48]. L. Xu, J. Qin, D. Sun, Y. Liao and J. Zheng, "PFformer: A Time-Series Forecasting Model for Short-Term Precipitation Forecasting," in IEEE Access, vol. 12, pp. 130948-130961, 2024.

[49]. A. Bojesomo, H. AlMarzouqi and P. Liatsis, "A Novel Transformer Network With Shifted Window Cross-Attention for Spatiotemporal Weather Forecasting," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 45-55, 2024.

[50]. G. Xu et al., "TLS-MWP: A Tensor-Based Long- and Short-Range Convolution for Multiple Weather Prediction," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 9, pp. 8382-8397, Sept. 2024.

[51]. S. Kumari, P. Muthulakshmi, "SARIMA Model: An Efficient Machine Learning Technique for Weather Forecasting", Procedia Computer Science, Vol. 235, pp. 656-670, 2024.

[52]. M. Ciszynski, K. Chrominski, "Applying machine learning techniques in weather forecast modeling", Procedia Computer Science, Vol. 246, pp. 4133-4141, 2024.

[53]. D. Pantola, Ma. Gupta, M. Agarwal, R. Bohra, K. Rawat, "Predictive Analytics in Weather Forecasting Using Machine Learning and Deep Learning", Artificial Intelligence and Knowledge Processing, AIKP 2024, Communications in Computer and Information Science, vol 2228, pp. 103-116, 2024.

[54]. O. Sharma, D. Trivedi, S. Pattnaik, V. Hazra and N. B. Puhan, "Improvement in District Scale Heavy Rainfall Prediction Over Complex Terrain of North East India Using Deep Learning," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-8, 2023.

[55]. M. Choubisa, M. Dubey, N. Jain and H. Virwani, "Systematic Analysis of Weather Prediction for Jaipur City Dataset Using Deep Learning," 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), pp. 1-5, 2023.

[56]. G. Suthar, S. Singh, N. Kaul, S. Khandelwal, R. P. Singhal, "Prediction of maximum air temperature for defining heat wave in Rajasthan and Karnataka states of India using machine learning approach", Remote Sensing Applications: Society and Environment, Vol. 32, 2023.

[57]. S. Verma, A. Agarwal and H. Sharma, "Predicting Rain and Thunderstorm in Jaipur Using Machine Learning Techniques," 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM), pp. 1-4, 2023.

[58]. M. M. Hassan et al., "Machine Learning-Based Rainfall Prediction: Unveiling Insights and Forecasting for Improved Preparedness," in IEEE Access, vol. 11, pp. 132196-132222, 2023.

[59]. Ryoya Tanabe, Tsutomu Matsui, Takashi S.T. Tanaka, "Winter wheat yield prediction using convolutional neural networks and UAV-based multispectral imagery", Field Crops Research, Vol. 291, February 2023.

[60]. P. R. Gangula, J. Yeboah and I. K. Nti, "A Comparative Study of Machine Learning Techniques for Nuanced Weather Prediction," 2023 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 260-265, 2023.

[61]. S. Iram et al., "An Innovative Machine Learning Technique for the Prediction of Weather Based Smart Home Energy Consumption," in IEEE Access, vol. 11, pp. 76300-76320, 2023.

[62]. Liuyi Chen,Bocheng Han,Xuesong Wang, Jiazhen Zhao, Wenke Yang and Zhengyi Yang, "Machine Learning Methods in Weather and Climate Applications: A Survey", Applied sciences, Special Issue Methods and Applications of Data Management and Analytics, pp. 1-36, 2023.

[63]. P. S. Mung and S. Phyu, "Time Series Weather Data Forecasting Using Deep Learning," IEEE Conference on Computer Applications (ICCA), pp. 254-259, 2023.

[64]. G. Zenkner, S. N. Martinez, "A flexible and lightweight deep learning weather forecasting model", Applied Intelligence, Vol. 53, pp. 24991–25002, 2023.

[65]. H. Choi, J. Choi, J. Hwang, K. Lee, D. Lee, N. Park, "Climate modeling with neural advection–diffusion equation", Knowledge and Information Systems, Vol. 65, pp. 2403-2427, 2023.

[66]. F Najib and I W Mustika, "Weather Forecasting Using Artificial Neural Network for Rice Farming in Delanggu Village", IOP Conference Series: Earth and Environmental Science, pp. 1-9, 2022.

[67]. Tulsi Pawan Fowdur, Rosun Mohammad Nassir-Ud-Diin Ibn Nazir, "A real-time collaborative machine learning based weather forecasting system with multiple predictor locations", Array, Vol. 14, pp. 1-13, 2022.

[68]. F. R. Aderyani, S. J. Mousavi, F. Jafari, "Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN", Journal of Hydrology, Vol. 614, Part A, 2022.

[69]. R. Meenal, K. Kailash, P. A. Michael, J. J. Joseph, F. T. Josh, E. Rajasekaran, "Machine learning based smart weather prediction", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 28, No. 1, pp. 508-515, 2022.

[70]. Bogdan Bochenek, Zbigniew Ustrnul, "Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives", Atmosphere, Special Issue Artificial Intelligence in Atmospheric Modelling, Prediction, and Data Assimilation, pp. 1-16, 2022.

[71]. S. Gebretinsae, M. W. Ahmad, T.Gopi Krishna, N. Sreenivas, B. Yirga, "Machine Learning-Based Intelligent Weather Monitoring and Predicting System", International Journal of Engineering Trends and Technology, Vol. 70, Issue. 4, pp. 152-163, 2022.

[72]. J. Faraji, A. Ketabi, H. Hashemi-Dezaki, M. Shafie-Khah and J. P. S. Catalão, "Optimal Day-Ahead Self-Scheduling and Operation of Prosumer Microgrids Using Hybrid Machine Learning-Based Weather and Load Forecasting," in IEEE Access, vol. 8, pp. 157284-157305, 2020.

[73]. M. S. Hossain and H. Mahmood, "Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network and Synthetic Weather Forecast," in IEEE Access, vol. 8, pp. 172524-172533, 2020.

[74]. C. Zhang et al., "Weather Visibility Prediction Based on Multimodal Fusion," in IEEE Access, vol. 7, pp. 74776-74786, 2019.

[75]. S. Singh, M. Kaushik, A. Gupta, A. K. Malviya, "Weather Forecasting Using Machine Learning Techniques", Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) 2019, pp. 1-6, 2019.

[76]. J. L. Aznarte and N. Siebert, "Dynamic Line Rating Using Numerical Weather Predictions and Machine Learning: A Case Study," in IEEE Transactions on Power Delivery, vol. 32, no. 1, pp. 335-343, Feb. 2017.

[77]. S. K. Sagar, M. Rajeevan, S. V. B. Rao, A.K. Mitra, "Prediction skill of rainstorm events over India in the TIGGE weather prediction models", Atmospheric Research, Vol. 198, pp. 194-204, 2017.